JOURNAL OF APPLIED BEHAVIOR ANALYSIS 2012, **45,** 329–344 NUMBER 2 (SUMMER 2012)

AN EVALUATION OF THE STIMULUS EQUIVALENCE PARADIGM TO TEACH SINGLE-SUBJECT DESIGN TO DISTANCE EDUCATION STUDENTS VIA BLACKBOARD

BROOKE D. WALKER AND RUTH ANNE REHFELDT

SOUTHERN ILLINOIS UNIVERSITY

The purpose of the current study was to examine the degree to which instruction based on stimulus equivalence procedures could be used to teach single-subject design methodology to graduate-level professionals through a Web-based course management system known as Blackboard (see http://www.blackboard.com). Specifically, we used the stimulus equivalence paradigm to teach relations among the names, definitions, graphical representations of the designs, and two practical scenarios of when it would be appropriate to implement each design. Most participants demonstrated the emergence of untaught relations, and some participants showed generalization to novel vignettes and graphs. Relations largely were not maintained at follow-up but were retaught.

Key words: distance education, college teaching, derived stimulus relations, verbal behavior

---

Unique educational technologies, such as digital media and Web-based learning, have drastically changed the teaching and learning environment in higher education (Black & Watties-Daniels, 2006). Blackboard is a third-party Web-based course development platform that allows instructors to post information, documents, assignments, and announcements. Instructors and students also can engage in synchronous activities in chat rooms and in asynchronous activities such as discussion boards and private communication between the instructor and student (Servonsky, Daniels, & Davis, 2005). Liaw (2008) reported a number of benefits of Blackboard and other learning management systems for distance education students, including the fact that students can access the learning program at their convenience, students and instructors do not have to meet, and innovative approaches to teaching become available at an economically feasible cost (see also Capper, 2001). Course management systems such as Blackboard undoubtedly can play a vital role in behavior analysis instruction, but, as noted by Osborn (2010), instructional resources must be evaluated before being judged useful.

The derived stimulus relations instructional framework is a research framework that could be incorporated easily into an online course management system, such as Blackboard, for students enrolled in distance education. Recently, the paradigm has been implemented successfully as a curricular tool in higher education. For example, Ninness et al. (e.g., 2005, 2006, 2009) used an automated procedure to teach complex mathematical relations to college students who lacked formal instruction in algebraic and trigonometric transformation. Participants were taught to conditionally relate standard to factored formulas and factored formulas to graphs. Most participants demonstrated the emergence of untaught relations between the stimuli, and responding generalized to include a number of complex variations of the original instructional formulas and graphs (Ninness et al., 2005; see also Ninness et al., 2006, 2009). Likewise, Fields et al. (2009) focused on the concept of statistical interaction effects in two-way analyses of variance with undergraduate psychology students. Stimuli included graphical representations, written descriptions, and names and definitions for each interaction effect (e.g., crossover, synergistic,

---

Address correspondence to Ruth Anne Rehfeldt, Rehabilitation Institute, Behavior Analysis and Therapy Program, Mail Code 4609, Southern Illinois University, Carbondale, Illinois 62901 (e-mail: rehfeldt@siu.edu).

doi: 10.1901/jaba.2012.45-329

divergent, and no interaction). Following mastery of the automated teaching protocol, most participants demonstrated class-consistent responding that generalized to novel stimuli and novel questions presented in a paper-and-pencil test. Moreover, participants performed better on the paper-and-pencil test than a group of participants who did not complete the protocol. Finally, Fienup, Covey, and Critchfield (2010) established relations between anatomical location and functions of brain regions as a laboratory exercise with introductory psychology students. Together, these studies illustrate how both online and on-campus courses can make use of the derived stimulus relations framework either as a course requirement or as an out-of-class laboratory assignment. In fact, this approach seems to capture a number of best practices in instruction articulated by Skinner (2003), including frequent opportunities for feedback, ongoing evaluation, and the requirement that students master one set of skills before advancing to the next (see Critchfield & Fienup, 2008). Even more important, such procedures appear to engender skills above and beyond those that have been taught explicitly, or, as described by Critchfield and Fienup (2008, p. 363) "spawn novel abilities," which is a goal in any college classroom.

For the research program on derived stimulus relations to continue to develop as an instructional technology (see Rehfeldt, 2011; Rehfeldt & Barnes-Holmes, 2009), a promising approach may be to synthesize this research program with that based on Skinner's (1957) taxonomy of verbal operants, an approach that is presently popular as an instructional curriculum for learners with disabilities (e.g., Sundberg, 2008). Incorporating Skinner's framework into the study of derived stimulus relations in the college classroom may provide additional information on the controlling conditions under which novel verbal repertoires emerge, as well as contribute to the understanding of the acquisition and generalization of more sophisticated verbal repertoires than typically have been targeted in investigations of Skinner's taxonomy (Dixon, Small, & Rosales, 2007).

The *intraverbal,* defined by Skinner as a vocal or written response under the control of a vocal or written stimulus with which it shares no point-to-point correspondence (1957, p. 71), was identified by Sautter and LeBlanc (2006) as encompassing both the most diverse and least investigated of Skinner's verbal operants. To date, much of the limited amount of research on intraverbal behavior has focused on the direct teaching of intraverbals (i.e., Finkel & Williams, 2001), granting little attention to the emergence of an untaught intraverbal repertoire following instruction in another operant repertoire (see Perez-Gonzalez, Herszlikowicz, & Williams, 2008). Intraverbal repertoires may be difficult to establish, given the lack of physical discriminative stimuli present to occasion correct responding (Skinner, 1957). Thus, there may be some utility to examining economical and efficient procedures, potentially captured by the derived stimulus relations framework, for promoting the emergence of intraverbal responding (Perez-Gonzalez et al., 2008). Walker, Rehfeldt, and Ninness (2010) recently exemplified these efforts in a college course by demonstrating the emergence of untaught topography-based intraverbals, both vocal and written, following the direct teaching of selection-based intraverbals using a paper-and-pencil instructional format developed by Eikeseth, Rosales-Ruiz, Duarte, and Baer (1997). (See also Lovett, Rehfeldt, Garcia, & Dunning, 2011, who showed the emergence of untaught tacts using a paper-and-pencil stimulus equivalence protocol designed for use with college students.)

Walker et al. (2010) prescribed a number of new directions in the further development of the derived stimulus relations framework as a curricular tool in higher education. First, all of the studies published to date have included undergraduates as research participants. Focusing on more diverse students, such as graduate students or students who take online courses, also might be of value. In fact, given the wide accessibility of Blackboard technology in many university settings, automated protocols could be incorporated easily as class assignments in online courses. Second, additional studies that examine the generalization of derived stimulus relations to novel exemplars of original instructional stimuli are in order (e.g., Ninness et al. 2005, 2006, 2009). Third, few applied studies have examined the long-term maintenance of derived stimulus relations (see Lane & Critchfield, 1998; Ramirez & Rehfeldt, 2009; Rosales & Rehfeldt, 2007; Walker et al., 2010). Undoubtedly, the length of time knowledge sets are retained by students is important to all college instructors. Studies are needed that demonstrate the long-term stability of emergent relations. In addition, if the relations are not retained, how much reteaching of the baseline relations is necessary to recapture the emergent performances?

The purpose of the present study was to explore the efficacy of the stimulus equivalence paradigm for teaching single-subject methodology to graduate-level professionals enrolled in an online behavior analysis course. In the present study, we established selection-based intraverbals and evaluated the emergence of written topography-based intraverbals and generalization of the relations to novel graphs and clinical vignettes. Many academic skills, such as being able to speak and write about one's subject, involve topography-based responding, so we particularly were interested in determining whether our instructional procedures would be sufficient for producing emergent topography-based intraverbals. Participants were tested for the stability of the emergent relations 16 weeks after their initial completion of the protocol. If participants did not maintain the relations at follow-up, they received remedial instructional sessions to reestablish baseline relations, and then emergent relations were retested. The entire instructional and test protocol was presented online via Blackboard.

METHOD

Participants

Eleven graduate-level professionals, enrolled in an online course on behavioral assessment and observation methods through Southern Illinois University's Behavior Analysis and Therapy online program, participated in the study during the Spring 2010 semester. All of the 36 students in the course were living in different locations throughout the United States and enlisted in the class to fulfill a requirement for taking the national examination to become Board Certified Behavior Analysts (BCBA). Of the 11 students who participated in the study, seven completed maintenance probes and remedial teaching at the end of the course. Students were compensated with course credit for participating and provided informed consent for their participation.

Setting and Stimulus Materials

Experimental sessions lasted approximately 60 to 90 min and took place on the Blackboard course server through which the course was delivered. The stimuli were identical to those used in Lovett et al. (2011), and included names, definitions, graphical representations, and clinical vignettes of withdrawal, multiple baseline, alternating treatment, and changing criterion single-subject designs. The design names (A stimuli) and definitions (B stimuli) were obtained from Kennedy (2004), a book frequently used in the teaching of single-subject research methodology. Graphical representations and Graph Variant 1 stimuli (graphical variations of the different designs) were constructed via Microsoft Excel using hypothetical data. Clinical vignettes (D stimuli) consisted of written applied scenarios for which use of a particular design would be appropriate, also based on Kennedy (2004), and text and vignette variation (D′1 and D′2) stimuli were constructed by the first author of the study. Graph Variant 2 (C′2) consisted of graphs found in research studies published in the *Journal of Applied Behavior Analysis.* Graph Variant 1 differed slightly from the original C stimuli in that the data trends were in opposite directions (see also Lovett et al.). Stimuli were designated with the following alphanumeric symbols: Names of designs (A stimuli); definitions of designs (B stimuli); graphical representations of designs (C stimuli); and clinical vignettes for which a particular design would be used to explore the efficacy of a given intervention (D stimuli). The combination of the A, B, C, and D stimuli constituted a class. The computerized, selection-based intraverbal instructional protocol was thus intended to result in the formation of four four-member stimulus classes including topography-based (i.e., typed responses) and selection-based (i.e., selecting a response option via a mouse click) intraverbal responding, with each class pertaining to a particular design, although we did not test all of the relations necessary to infer class formation. Rather, we evaluated only the tested relations that we believed to encompass clinically applicable skills (e.g., we did not test participants' relating of vignettes to graphs).

Test trials were comprised of three sets of 12 to 36 questions (B-A blocks or sets of trials included 12 questions, C-A blocks or sets of trials included 36 questions, D-A blocks or sets of trials included 36 questions, and D-B blocks or sets of trials included 36 questions) with definitions, graphical representations, or clinical vignettes displayed as the sample stimuli. One question at a time was presented on each test trial, with an answer box directly below it in which participants typed in their responses.

Design

A pretest–train–posttest–maintenance test sequence was used. Because the pretest–train–posttest sequence of the design was completed in one session that lasted approximately 60 to 90 min, threats to internal validity (e.g., history and maturation) were unlikely. Maintenance probes were completed 16 weeks after the original experiment for seven participants. Only seven of the original students were either still enrolled in the course at this time or were willing to complete maintenance probes.

Dependent Measure and Interobserver Agreement

The primary dependent measure was the percentage of correct responses during derived relations pre- and posttest, generalization, and maintenance probes. A derived stimulus relation was defined as accurate performance on 11 of 12 trials (92% correct) for each particular derived relation, including definition-to-name (B-A), graph-to-name (C-A), vignette-to-name (D-A), and vignette-to-definition (D-B) relations. A secondary measure was the number of 12-trial blocks required for participants to attain mastery criterion during the instructional protocols.

Interobserver agreement was recorded by an independent observer for 36% of all participants' pre- and posttest probe trials. The second observer viewed pre- and posttest probe data via printed copies of participants' tests. Interobserver agreement was calculated by dividing agreements plus disagreements for each item and multiplying by 100%. Resulting interobserver agreement was 96%. Scoring of participants' instructional trials was automated, so interobserver agreement data were not collected.

General Procedure

There were eight separate sections of the experiment (Figure 1), which participants were required to complete in sequential order. Section 1 included all topography-based tact (C-A, C′1-A, C′2-A) and intraverbal pretest probes (B-A, D-A, D′1-A, D′2-A, D-B, D′1-B, and D′2-B); Section 2 included selection-based intraverbal name-to-definition (A-B) instruction; Section 3 included topography-based intraverbal definition-to-name (B-A) posttest probes; Section 4 included selection-based intraverbal name-to-graph (A-C) instruction;

Section 5 included topography-based tact graph-to-name (C-A) posttest and generalization probes (C′1-A and C′2-A); Section 6 included selection-based intraverbal definition-to-vignette (B-D) instruction; Section 7 included topography-based intraverbal vignette-to-name (D-A) posttest probes and generalization probes (D′1-A and D′2-A); and Section 8 included topography-based intraverbal vignette-to-definition (D-B) posttest and generalization probes (D′1-B and D′2-B). (The entire set of trials taught and tested is available from the corresponding author.) On all trials, participants were given 20 s to respond. This is the time limit the authors typically allow students to respond to test questions in online courses; it is enough time for students to provide a correct answer but not enough time for students to consult their notes or textbooks.

Topography-based tact and intraverbal pretests. Section 1 assessed whether participants would name the correct design when presented with its definitions, graphs, and clinical vignettes (B-A, C-A, C′1-A, C′2-A, D-A, D′1-A, and D′2-A relations) and define each design when given clinical vignettes for which each design would be appropriate (D-B, D′1-B, and D′2-B relations), all of which required typed responses. The B-A intraverbal relations were evaluated first, followed by the C-A tact relations, D-A intraverbal relations, and D-B intraverbal relations. Generalization probes were included within the probes for each type of relation. Testing for the B-A intraverbal relations consisted of 12 trials; and testing for C-A tact, D-A intraverbal, and D-B intraverbal relations, including generalization probes, each consisted of 36 trials, totaling 120 pretest trials.

Prior to pretests, participants were given the following instructions:

> In this section of the experiment, you will be asked 120 questions. You have 40 minutes to answer ALL OF THEM. This gives you 20 seconds to complete each question so if you have not answered a question after 20 seconds, I encourage you to move on to the next by clicking "save and view next" below each item. Always click "save and view next" rather than "next question" so that your answers are always saved. Any answers submitted AFTER the 40 minutes will NOT be accepted. You will also NOT be able to revisit previous questions throughout the test. Once you click "save and view next," you cannot go back to that question. After answering item 120 and all questions have been attempted, click the "Finish" button. Be aware, some items require fill-in-the-blank answers and others require short answers. Please do your best to complete ALL 120 questions within the 40 minutes WITHOUT using your textbook or other study materials. Don't worry; we have NOT covered this material in class so I do not expect you to know it. This is to test what you know as of RIGHT NOW so please be honest and do not use supplemental aids to help in answering the questions. How you do on this experiment will NOT negatively affect your grade in any way. You will however receive 10 extra credit points for your participation in the experiment. Good luck and have fun!

The presentation of the sample and comparison stimuli together marked the onset of each test trial. Pre- and posttest probes, which required typed responses from participants, were scored by the experimenter because Blackboard scoring was insensitive to misspelling of words or did not accept abbreviations (i.e., MBL, CCD, ATD) as correct responses. Scoring was evaluated using an answer key that was developed for purposes of the study.

Selection-based intraverbal instruction. Participants were taught four name-to-definition (A-B), name-to-graph (A-C), and definition-to-vignette (B-D) intraverbal relations each. During name-to-definition (A-B) and name-to-graph (A-C) intraverbal instruction, the A stimuli, or design names, were conceptualized as the sample stimuli (i.e., withdrawal design), and the B or C stimuli, or the definitions or graphs, were conceptualized as the comparison stimuli (i.e., the design involves evaluation of the effects of a treatment by implementing and then removing the treatment, when its removal does not present risks to the client). During definition-to-vignette (B-D) intraverbal instruction, the B stimuli, or definitions, were conceptualized as sample stimuli, while the D stimuli, or clinical vignettes, were conceptualized as comparison stimuli.

Prior to teaching, participants were given the following instructions:

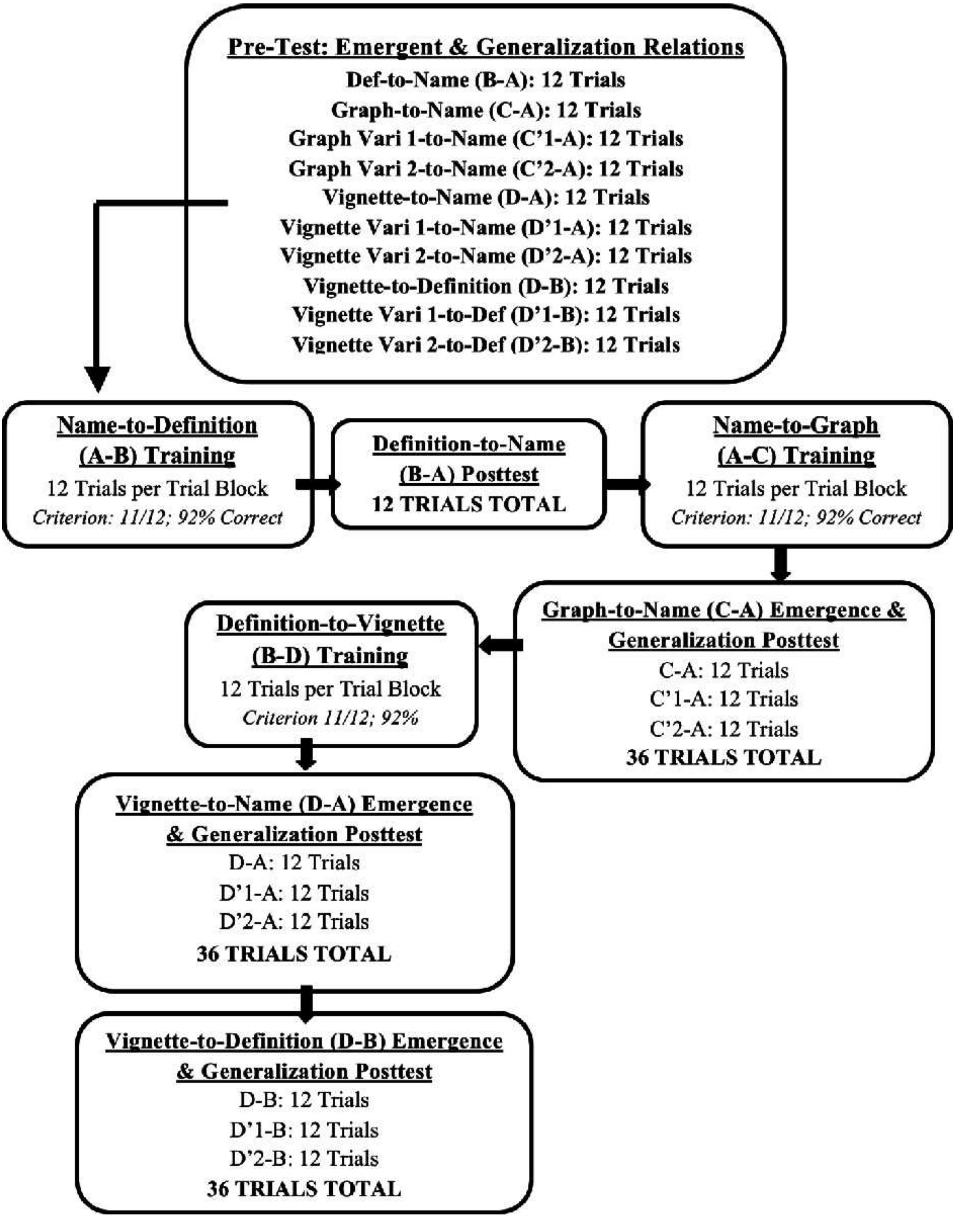

Figure 1. Flow chart of individual sections of the experiment.

> In this section of the experiment you will be asked to complete 12 multiple choice questions and you will have 6 minutes to do so. Once you have completed this section, your responses will be scored and Blackboard will show you your results. You must answer at least 11 out of 12 items correct before moving on to the following section of the experiment. If you do not meet this criterion, you will be asked to complete this section again until criteria are met. Once you have met this criterion, you may move on to the next section of the experiment.

Each teaching protocol consisted of 12 multiple-choice items, with the presentation of each item constituting a trial. The question in each trial was conceptualized as the sample stimulus, and the four response options were conceptualized as comparison stimuli. The sequence in which questions were presented was derived via a random sequence generator incorporated via Blackboard, as was the ordering of response options. Participants indicated their responses by clicking the mouse on the response option of their choice. They either selected the correct definition (A-B intraverbal instruction), graph (A-C intraverbal instruction), or clinical vignette (B-D intraverbal instruction) after presentation of the name (A-C and A-B instruction) or definition (B-D instruction) of the correct design. At the end of each trial block, participants were given feedback for the entire trial block for that section. Feedback was delivered in the form of "correct" or "incorrect" statements presented next to each item or trial in the block. A mastery criterion of 11 of 12 (92%) correct was required before a participant could advance to the next section. If a participant failed to attain mastery criterion, he or she was allowed time to look over his or her correct and incorrect answers, and then was asked to repeat the same section until mastery criterion was met. Examples of instructional trials are shown in Figure 2, which depicts an example of a name-to-definition (A-B) teaching trial, a name-to-graph (A-C) teaching trial, and a definition-to-vignette (B-D) teaching trial.

Topography-based tact and intraverbal posttests. Following each instructional section (Sections 2, 4, and 6), participants were administered topography-based posttest probes, which were intended to evaluate the emergence of derived topography-based intraverbal relations. Posttest probes were identical to pretest probes.

Topography-based tact and intraverbal response maintenance probes. During the last week of the course, 16 weeks after the initial experiment, all students who participated in the study were asked to complete an identical follow-up test to earn extra points toward their class grade. Maintenance probes were conducted to evaluate the stability of the emergent topography-based intraverbal and tact relations, and were identical to pre- and posttest probes.

Remedial instruction and posttests. Remedial selection-based intraverbal instruction, when needed, was conducted in a manner identical to that of the initial instruction, and posttest probes were presented again directly thereafter.

RESULTS

Pre- and posttest scores for topography-based tact and intraverbal test probes are shown in Figures 3 and 4. Figure 3 shows scores for untaught relations that involved stimuli presented during instruction, and Figure 4 depicts results for the same relations but with dimensional variants of instructional stimuli (i.e., generalization). All 11 participants scored in the range of 0% to 42% correct on pretest probes for all untaught relations. Table 1 indicates the number of teaching trial blocks required for each participant to meet the mastery criterion during the instruction of design name-to-definition (A-B), design name-to-graph (A-C), and definition-to-vignette (B-D) relations. Most participants required one or two trial blocks to master each taught relation.

For the posttest probes, the definition-to-name (B-A), graph-to-name (C-A), vignette-to-name (D-A), and vignette-to-definition (D-B) intraverbal relations as well as the generalized intraverbal relations (C′1-A, C′2-A, D′1-A, D′2-A, D′1-B, and D′2-B) were inferred to have emerged if a participant performed with 92% accuracy. As

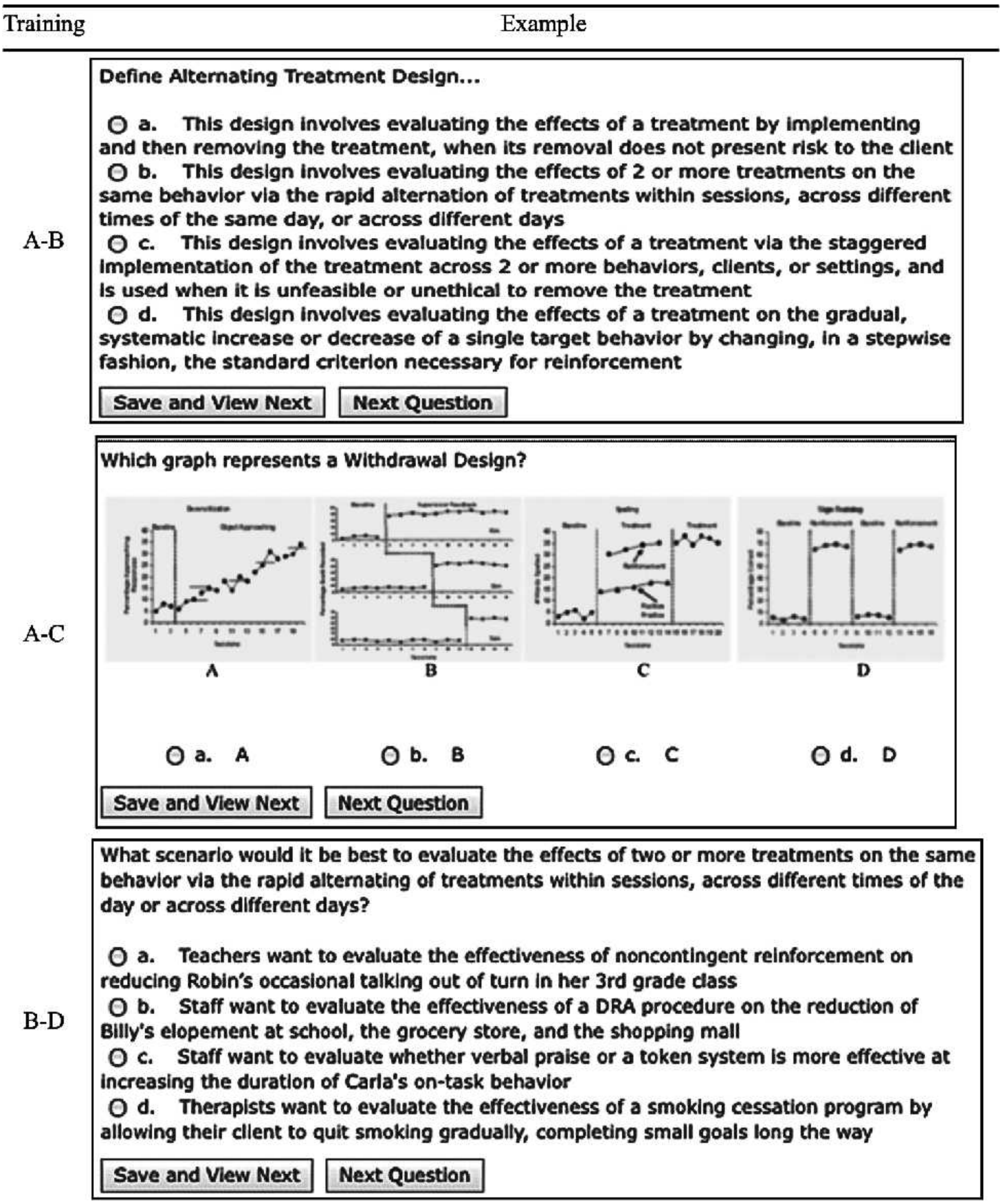

Figure 2. Examples of instructional trials.

shown in Figure 3, six participants demonstrated the emergence of the definition-to-name (B-A) intraverbal relations, all 11 demonstrated the emergence of the graph-to-name tact (C-A) relations, seven demonstrated the emergence of the vignette-to-name (D-A) intraverbal relations, and five demonstrated the emergence of the vignette-to-definition intraverbal (D-B) relations.

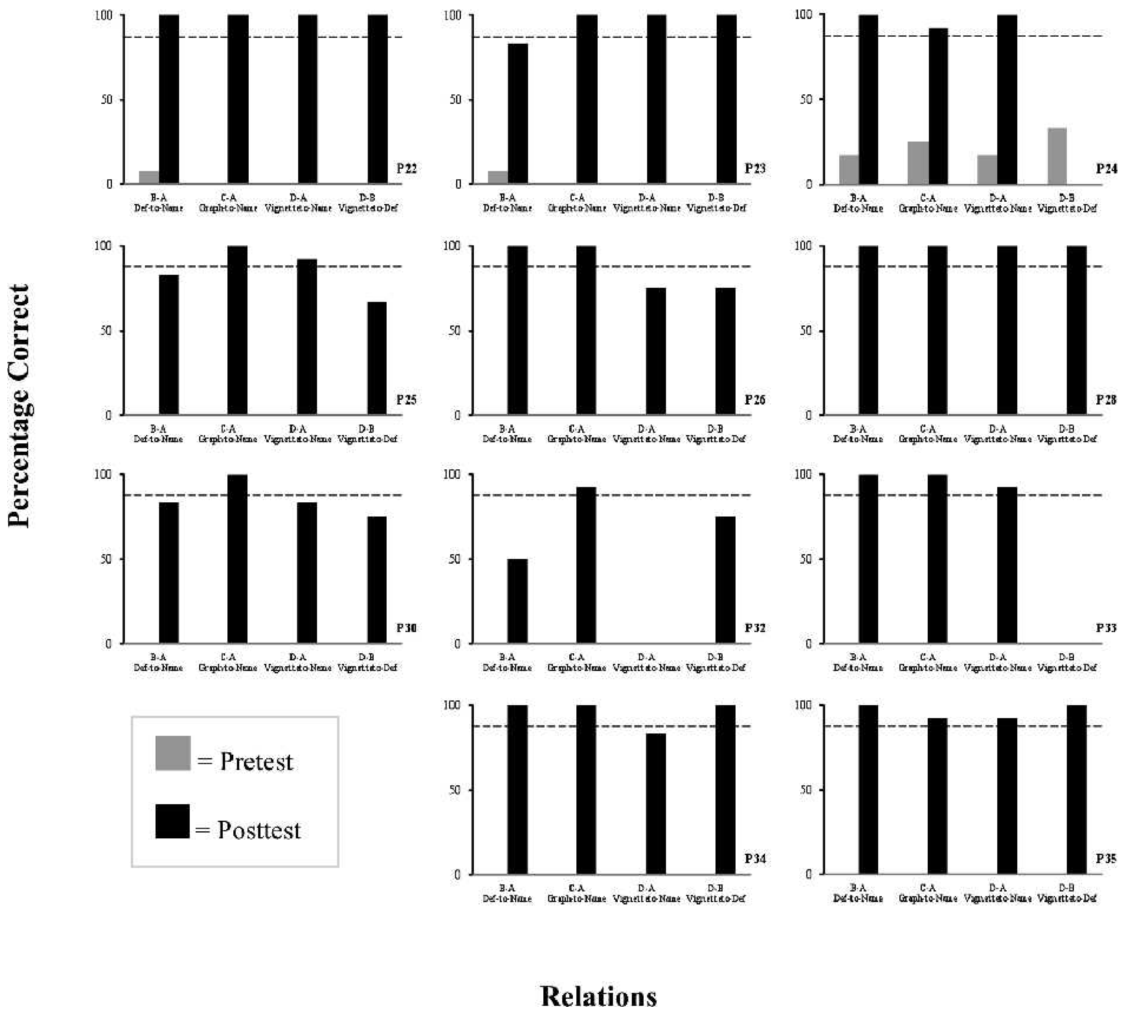

Figure 3. Percentage of correct responses during pre- and posttests for untaught relations. The dashed line represents mastery criterion.

Thus, all participants demonstrated the graph-to-name (C-A) tact relation most readily following selection-based name-to-graph (A-C) intraverbal instruction, whereas approximately half demonstrated the emergence of the definition-to-name (B-A), vignette-to-name (D-A), and vignette-to-definition (D-B) intraverbal relations.

As shown in Figure 4, nine participants showed generalization of the graph-to-name (C′1-A or C′2-A) tact relations, seven showed generalization of the vignette-to-name (D′1-A or D′2-A) intraverbal relations, and five showed generalization of the vignette-to-definition (D′1-B or D′2-B) intraverbal relations, to at least one novel variant of the C-A, D-A, and D-B relations.

Scores for pretest, posttest, and maintenance test probes for all emergent topography-based and generalized topography-based relations are shown in Figure 5. Few topography-based intraverbal relations were maintained 16 weeks after selection-based intraverbal instruction. None of the participants maintained the definition-to-name (B-A) intraverbal relations, and three participants scored above 50% correct, which was substantially higher than

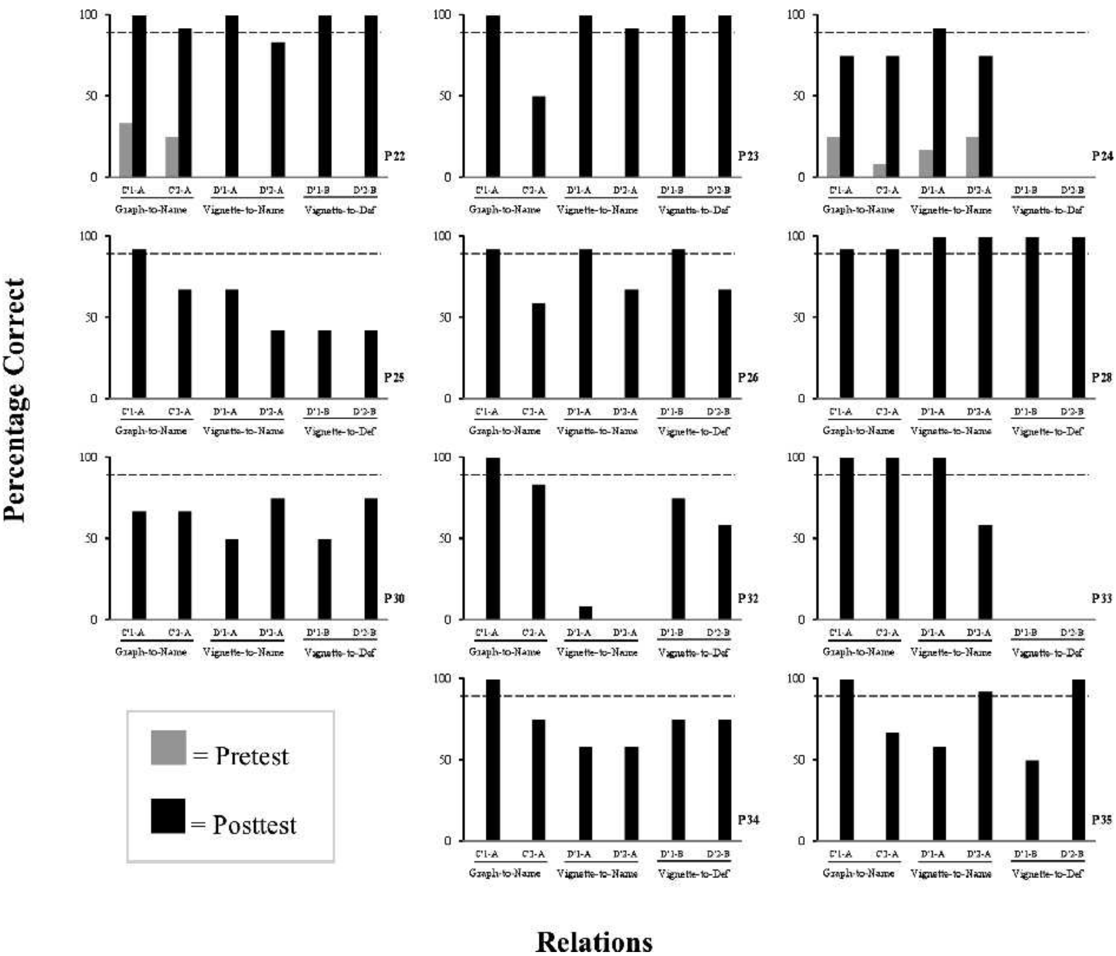

Figure 4. Percentage of correct responses during pre- and posttests for untaught relations with novel dimensional variants of instructional stimuli. The dashed line represents mastery criterion.

Table 1
Training Trial Blocks to Criterion During Initial Training Phases

Participant	Name to definition (A-B)	Name to graph (A-C)	Definition to vignette (B-D)
22	1	1	1
23	3	2	1
24	1	3	2
25	2	1	2
26	2	2	2
28	3	1	2
30	3	2	2
32	5	2	2
33	2	2	2
34	2	2	2
35	1	1	1

their initial pretest scores. Two participants maintained the graph-to-name (C-A) tact relations, and one participant scored above 50% correct. One participant maintained the vignette-to-name (D-A) intraverbal relations, and one participant scored above 50% correct. None of the participants maintained the vignette-to-definition (D-B) intraverbal relations, but three participants scored at or above 50% correct.

Figure 5 also shows that, of the seven participants who completed maintenance probes, less than half maintained the Graph Variant 1-to-name (C′1-A) generalized rela-

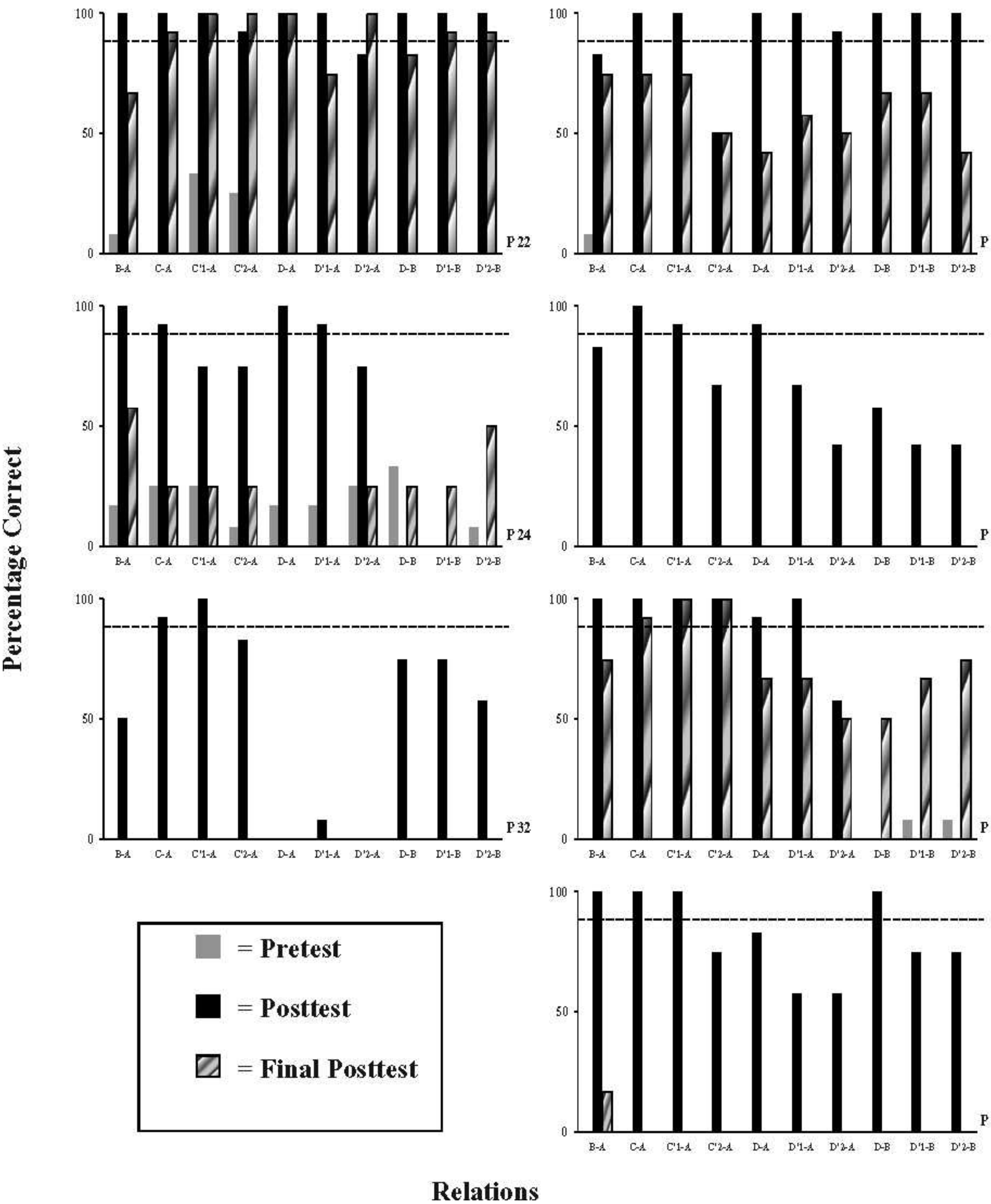

Figure 5. Percentage of correct responses during pretest, posttest, and maintenance test probes for all untaught relations. The dashed line represents mastery criterion.

tions; two participants maintained the Graph Variant 2-to-name (C′2-A) generalized relations; none of the participants maintained the Vignette Variant 1-to-name (D′1-A) relations; and one participant demonstrated maintenance of the Vignette Variant 2-to-name (D′2-A) relations. One participant demonstrated maintenance of generalized Vignette Variant 1-to-

Table 2
Training Trial Blocks to Criterion During Remedial Stability Training Phases

Participant	Name to definition (A-B)	Name to graph (A-C)	Definition to vignette (B-D)
22	1	1	1
23	1	1	3
24	1	1	1
25	2	3	3
32	4	2	2
33	1	1	2
34	2	2	1

definition (D′1-B) relations, and one participant demonstrated maintenance of the Vignette Variant 2-to-definition (D′2-B) generalized relations. Thus, most of the relations that involved novel dimensional variants of instructional stimuli were not intact for the participants 16 weeks after instruction.

Table 2 shows the number of teaching trial blocks required for each participant to reattain mastery criterion during remedial selection-based intraverbal instruction of the name-to-definition (A-B), name-to-graph (A-C), and definition-to-vignette (B-D) relations. As in the initial instruction, participants required approximately one to three teaching trial blocks to reattain mastery criterion for all of the relations. Figure 6 shows the first and second posttest probes following remedial instruction for emergent topography-based intraverbal relations, and Figure 7 shows the first and second posttest probes for generalized topography-based intraverbal relations. All seven of the participants who completed remedial instruction showed the reemergence of the definition-to-name (B-A) relations, three demonstrated the reemergence of the graph-to-name (C-A) relations, five demonstrated the reemergence of the vignette-to-name (D-A) relations, and three demonstrated the reemergence of the vignette-to-definition

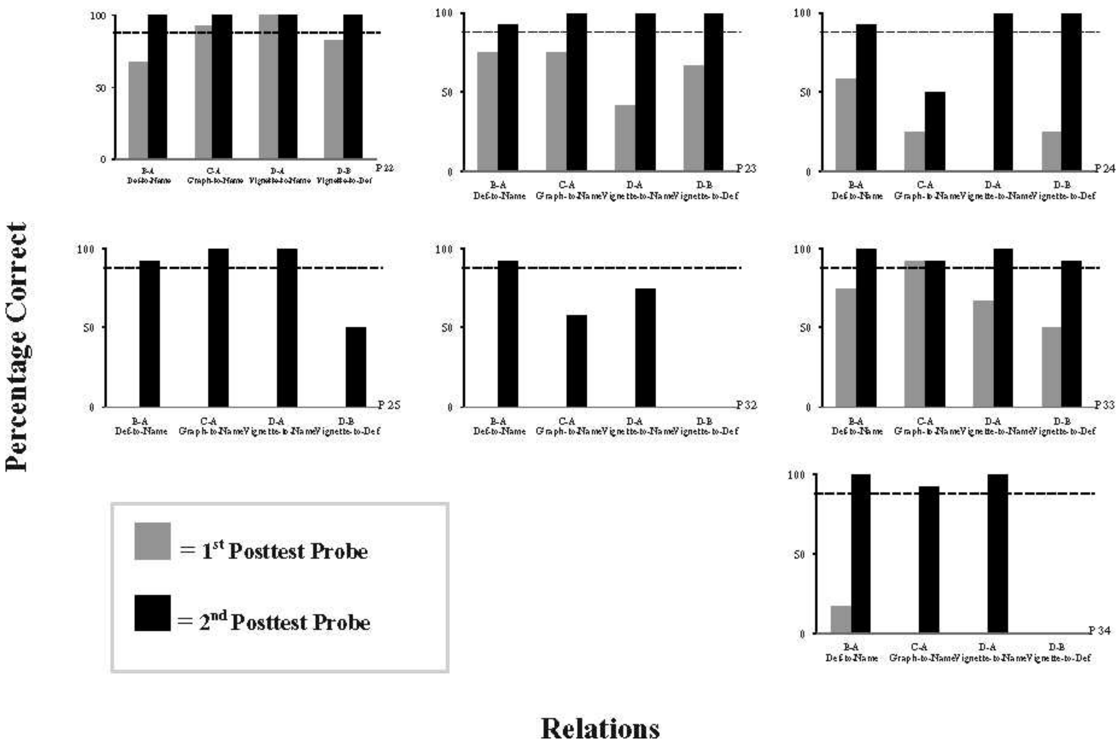

Figure 6. Percentage of correct responses on first and second posttest probes. The dashed line represents mastery criterion.

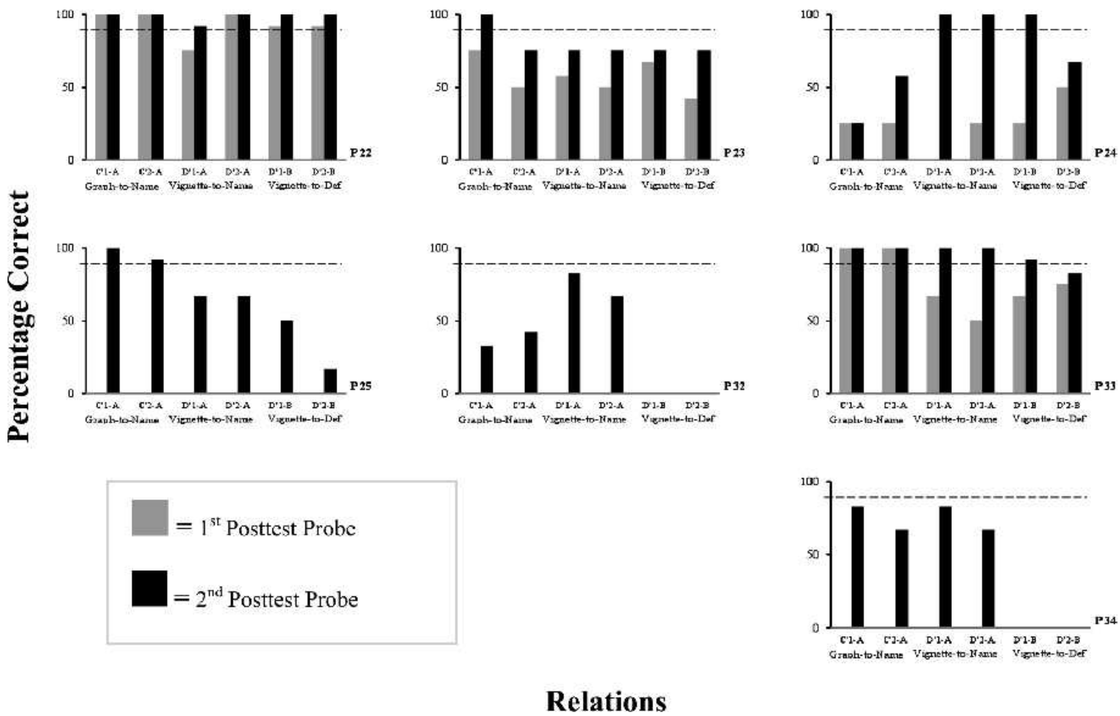

Figure 7. Percentage of correct responses on first and second posttest probes for the generalization relations. The dashed line represents mastery criterion.

(D-B) relations (Figure 6). Four showed reemerged generalization of the graph-to-name (C′1-A or C′2-A) tact relations, three showed reemerged generalization of the vignette-to-name (D′1-A or D′2-A) intraverbal relations, and three showed reemerged generalization of the vignette-to-definition (D′1-B or D′2-B) intraverbal relations. Generalization was observed to at least one novel variant of the C-A, D-A, and D-B relations (Figure 7). Thus, although emergent relations were not maintained at 16 weeks, remedial instruction was effective in reestablishing some of the relations, particularly the tact relations, for most of the participants. In addition, those who demonstrated emergent generalized topography-based relations during posttests did so after remedial instruction. Those who failed to demonstrate generalized relations during the posttests also failed to do so after the remedial instruction during posttests.

DISCUSSION

Our results demonstrate the utility of the derived stimulus relations framework, coupled with Skinner's (1957) analysis of verbal behavior, for promoting untaught skills in graduate students. The relatively simple selection-based intraverbal teaching protocol promoted the emergence of a number of novel topography-based responses. Although multiple-choice exams, which require selection-based responding, are widely used in many college courses, this repertoire is not necessarily the desired outcome after completion of a college course. Rather, written topography-based responses are more impressive demonstrations of emergent skills after effective teaching. These results thus join those of other studies that have illustrated the economy and efficiency of the stimulus equivalence paradigm in higher education by teaching broad, overarching concepts with minimal instruction (Fienup et al., 2010;

Ninness et al., 2005, 2006, 2009; Walker et al., 2010), and coincide with those of prior investigations that reported the emergence of untaught topography-based intraverbals following instruction in other verbal operant repertoires (see Chase, Johnson, & Sulzer-Azaroff, 1985; Perez-Gonzalez et al., 2008; Walker et al., 2010). It is noteworthy that these findings were obtained from a class that was taught exclusively online. The flexibility of Blackboard and other online course management software programs may permit the construction of any number of widely varied equivalence protocols that can be completed by many learners at any time. Given the challenges distance education presents for effective teaching (Clerehan, 2003), protocols that "spawn novel abilities" (Critchfield & Fienup, 2008, p. 363) are welcome additions to online courses.

The present results also lend support to the utility of the paper-and-pencil instructional protocol developed by Eikeseth et al. (1997), who strongly endorsed the face validity of this teaching approach over that of the more traditional match-to-sample arrangement, which, as noted by Eikeseth et al., was developed for research with pigeons (i.e., Cumming & Berryman, 1961). Paper-and-pencil teaching protocols are not unlike multiple-choice questionnaires widely used in classroom settings, and our results illustrate how the protocol established by Eikeseth et al. could be created easily for online learning. Not only was the teaching protocol administered to students at a distance, but it was also efficient from both a learning and instructional standpoint; most participants mastered the selection-based intraverbal relations in few teaching blocks, and, aside from the initial programming, little effort was required on the part of the experimenter. Not surprisingly, however, untaught tacts were shown to emerge more readily than untaught intraverbals, a necessarily more complex verbal operant that is under the control of verbal, as opposed to physical, discriminative stimuli (Sautter & LeBlanc, 2006; Skinner, 1957). In our procedure, graph-to-name (C-A) tact test trials consisted of the presentation of graphic or pictoral sample stimuli and required a brief typed response, whereas the vignette-to-name (D-A) and vignette-to-definition (D-B) intraverbal test trials consisted of textual stimuli as sample stimuli and required lengthier typed responses. Thus, there may be limits to the complexity of the response repertoires that simple selection-based paper-and-pencil teaching may engender. For such complex skills, multiple-exemplar instruction may be necessary, in which participants are taught explicitly all of the baseline and emergent relations with specific sets of stimuli before emergent responses can occur with novel sets (see Berens & Hayes, 2007). Explicit instruction in both selection- and topography-based repertoires may be essential.

Generalization results for this experiment were somewhat disappointing, because not all of the emergent relations were shown to generalize to novel stimuli for all of the participants. Indeed, generalization of knowledge learned with one set of stimuli to novel exemplars seems to be among the most important outcomes for this experiment, and, in fact, among the most important outcomes for a graduate-level clinician completing a course on this topic. These generalization failures point to the possibility that participants' responding may have been under the control of irrelevant features of the stimuli presented during instruction, such as single words or the sequence of words. For example, the multiple baseline design vignettes could be identified simply by the presence of a list of three things, and the word "goal" appeared in the vignettes for changing criterion designs. Teaching with multiple exemplars of stimuli, such as definitions and vignettes that were worded differently, may have circumvented this problem.

Maintenance results also merit mention, because only two participants demonstrated criterion performance on maintenance test probes in the absence of any remedial instruction. These results are not surprising, given that not all of the participants demonstrated criterion performance on all posttest probes during initial posttests. Few investigations of derived stimulus relations have examined the long-term stability of the emergent relations (Rehfeldt & Hayes, 2000; Saunders, Wachter, & Spradlin, 1988; Walker et al., 2010), and the identification of protocols that engender long-term maintenance is critical for ascertaining the applied value of derived stimulus relations protocols. Our results suggest that emergent relations may be retained over time in the absence of instruction, and, when necessary, only a small amount of reteaching is necessary to recapture them, in that most participants required only one or two remedial instruction blocks to demonstrate criterion performance again.

In addition, our examination of students' grade point averages and course performance revealed that the majority of participants who demonstrated high levels of accuracy during posttests hold a grade point average above 3.0. Continued efforts to explore the relation between academic skill levels and performance on protocols such as these are in order. For example, O'Hora, Pelaez, and Barnes-Holmes (2005) established correlations between performance on relational tasks and vocabulary and arithmetic subtests of the Wechsler Adult Intelligence Scale (Wechsler, 1997). Whether a direct relation exists between performance on protocols such as ours and academic performance remains to be seen.

The pretest–train–posttest design sequence used in this study limits the firmness of conclusions that can be drawn from the results. Although it is unlikely that extraneous variables were responsible for changes in participants' performance from pretest to posttest, the efficacy of our protocol would have been underscored had participants' performance been compared to that of another group of participants who participated in a more standard form of instruction, such as a lecture on single-subject design. Indeed, this was the focus of Lovett et al. (2011), who also used single-subject design concepts to illustrate the utility of the derived stimulus relations framework in higher education. In Lovett et al., however, an automated match-to-sample protocol was found to be as equally effective as a lecture in enhancing performance on a short exam, and, moreover, was not found to be more preferred by students than the lecture. Future comparisons of equivalence protocols to more traditional educational protocols are in order.

In conclusion, these results pave the way for an incorporation of the derived stimulus relations teaching technology into distance education. The results also shed light on the transfer of skills from one verbal operant repertoire to another. Future research should further test the limits of selection-based instruction, as well as examine the role that teaching with multiple exemplars may have in promoting the emergence of a range of untaught skills. Overall, the instructional protocol employed in the present study seems to be highly consistent with Skinner's (2003) notion of best practices in instructional technology.

REFERENCES

Berens, N. M., & Hayes, S. C. (2007). Arbitrarily applicable comparative relations: Experimental evidence for relational operant. *Journal of Applied Behavior Analysis*, *40*, 45–71.

Black, C. D., & Watties-Daniels, A. D. (2006). Cutting edge technology to enhance nursing classroom instruction at Copin State University. *Official Journal of the Association of Black Nursing Faculty in Higher Education*, *17*, 103–106.

Capper, J. (2001). *E-learning growth and promise for the developing world*. Retrieved from http://www.techknowlogia.org

Chase, P. N., Johnson, K. R., & Sulzer-Azaroff, B. (1985). Verbal relations within instruction: Are there subclasses of the intraverbal? *Journal of the Experimental Analysis of Behavior*, *43*, 301–313.

Clerehan, R. (2003). Transition to tertiary education in the arts and humanities: Some academic initiatives from Australia. *Arts and Humanities in Higher Education*, *2*, 72–89.

Critchfield, T. S., & Fienup, D. M. (2008). Stimulus equivalence. In S. F. Davis & W. F. Buskist (Eds.), *21st century psychology* (pp. 360–372). Thousand Oaks, CA: Sage.

Cumming, W. W., & Berryman, R. (1961). Some data on matching behavior in the pigeon. *Journal of the Experimental Analysis of Behavior*, *4*, 281–284.

Dixon, M. R., Small, S. L., & Rosales, R. (2007). Extended analysis of empirical citations with Skinner's *Verbal Behavior*: 1984–2004. *The Behavior Analyst*, *30*, 197–209.

Eikeseth, S., Rosales-Ruiz, J., Duarte, A., & Baer, D. M. (1997). The quick development of equivalence classes in a paper-and-pencil format through written instructions. *The Psychological Record*, *47*, 275–284.

Fields, L., Travis, R., Yadlovker, E. D., Roy, D., Aguiar-Rocha, L., & Sturmey, P. (2009). Equivalence class formation: A method for teaching statistical interactions. *Journal of Applied Behavior Analysis*, *42*, 595–608.

Fienup, D. M., Covey, D. P., & Critchfield, T. S. (2010). Teaching brain-behavior relations economically with stimulus equivalence technology. *Journal of Applied Behavior Analysis*, *43*, 19–33.

Finkel, A. S., & Williams, R. L. (2001). A comparison of textual and echoic prompts on the acquisition of intraverbal behavior in a six-year-old boy with autism. *The Analysis of Verbal Behavior*, *18*, 61–70.

Kennedy, C. (2004). *Single-case designs for educational research*. Boston: Allyn and Bacon.

Lane, S. D., & Critchfield, T. S. (1998). Classification of vowels and consonants by individuals with moderate mental retardation: Development of arbitrary relations via match-to-sample training with compound stimuli. *Journal of Applied Behavior Analysis*, *31*, 21–41.

Liaw, S. (2008). Investigating students' perceived satisfaction, behavioral intention, and effectiveness of e-learning: A case study of the Blackboard system. *Computers & Education*, *51*, 864–873.

Lovett, S., Rehfeldt, R., Garcia, Y., & Dunning, J. (2011). Comparison of a stimulus equivalence protocol and traditional lecture for teaching single-subject designs. *Journal of Applied Behavior Analysis*, *44*, 819–833.

Ninness, C., Barnes-Holmes, D., Rumph, R., McCuller, G., Ford, A. M., Payne, R., et al. (2006). Transformations of mathematical and stimulus functions. *Journal of Applied Behavior Analysis*, *39*, 299–321.

Ninness, C., Dixon, M., Barnes-Holmes, D., Rehfeldt, R. A., Rumph, R., McCuller, G., et al. (2009). Constructing and deriving reciprocal trigonometric relations: A functional analytic approach. *Journal of Applied Behavior Analysis*, *42*, 191–208.

Ninness, C., Rumph, R., McCuller, G., Harrison, C., Ford, A. M., & Ninness, S. K. (2005). A functional analytic approach to computer-interactive mathematics. *Journal of Applied Behavior Analysis*, *38*, 1–22.

O'Hora, D., Pelaez, M., & Barnes-Holmes, D. (2005). Derived relational responding and performance on verbal subtests of the WAIS-III. *The Psychological Record*, *55*, 155–175.

Osborn, D. R. (2010). Do print, web-based, or blackboard integrated tutorial strategies differentially influence student learning in an introductory psychology class? *Journal of Instructional Psychology*, *3*, 247–251.

Perez-Gonzalez, L. A., Herszlikowicz, K., & Williams, G. (2008). Stimulus relations analysis and the emergence of novel intraverbals. *The Psychological Record*, *58*, 95–129.

Ramirez, J., & Rehfeldt, R. A. (2009). Observational learning and the emergence of symmetry relations in teaching Spanish vocabulary words to typically developing children. *Journal of Applied Behavior Analysis*, *42*, 801–805.

Rehfeldt, R. A. (2011). Toward a technology of derived stimulus relations: An analysis of articles published in the *Journal of Applied Behavior Analysis,* 1992–2009. *Journal of Applied Behavior Analysis*, *44*, 109–119.

Rehfeldt, R. A., & Barnes-Holmes, Y. (2009). *Derived relational responding applications for learners with autism and other developmental disabilities*. Oakland, CA: New Harbinger.

Rehfeldt, R. A., & Hayes, L. J. (2000). The long-term retention of generalized equivalence classes. *The Psychological Record*, *50*, 405–428.

Rosales, R., & Rehfeldt, R. A. (2007). Contriving transitive conditioned establishing operations to establish derived manding skills in adults with severe developmental disabilities. *Journal of Applied Behavior Analysis*, *40*, 105–121.

Saunders, R. R., Wachter, J., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior*, *59*, 95–115.

Sautter, R. A., & LeBlanc, L. A. (2006). The empirical applications of Skinner's analysis of verbal behavior with humans. *The Analysis of Verbal Behavior*, *22*, 35–48.

Servonsky, E. J., Daniels, W. L., & Davis, B. L. (2005). Evaluation of Blackboard as a platform for distance education. *The ABNF Journal*, *16*(6), 132–135.

Skinner, B. F. (1957). *Verbal behavior*. Englewood Cliffs, NJ: Prentice Hall.

Skinner, B. F. (2003). *The technology of teaching*. Acton, MA: Copley.

Sundberg, M. L. (2008). *The verbal behavior milestones assessment and placement program.* Concord, CA: AVB Press.

Walker, B., Rehfeldt, R. A., & Ninness, C. F. (2010). Using the stimulus equivalence paradigm to teach course material in an undergraduate rehabilitation course. *Journal of Applied Behavior Analysis*, *43*, 615–633.

Wechsler, D. (1997). *Wechsler adult intelligence scale* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Received October 1, 2010
Final acceptance September 21, 2011
Action Editor, Thomas Higbee